Multiple Source Spatial Cluster Detection Through Multi-criteria Analysis

ISDS 2012 Conference Abstracts

Multiple Source Spatial Cluster Detection Through Multi-criteria Analysis

Luiz H. Duczmal*[1], Alexandre C. L. Almeida[2], Fabio R. da Silva[1] and Martin Kulldorff[3]

[1]Universidade Federal de Minas Gerais, Belo Horizonte, Brazil; [2]Universidade Federal de São João del-Rei, Ouro Branco, Brazil; [3]Harvard Medical School, Boston, MA, USA

Objective

To incorporate information from multiple data streams of disease surveillance to achieve more coherent spatial cluster detection using statistical tools from multi-criteria analysis.

Introduction

Multiple data sources are essential to provide reliable information regarding the emergence of potential health threats, compared to single source methods [1,2]. Spatial Scan Statistics have been adapted to analyze multivariate data sources [1]. In this context, only ad hoc procedures have been devised to address the problem of selecting the most likely cluster and computing its significance. A multi-objective scan was proposed to detect clusters for a single data source [3].

Methods

For simplicity, consider only two data streams. The j-th objective function evaluates the strength of candidate clusters using only information from the j-th data stream. The best cluster solutions are found by maximizing two objective functions simultaneously, based on the concept of dominance: a point is called dominated if it is worse than another point in at least one objective, while not being better than that point in any other objective [4]. The nondominated set consists of all solutions which are not dominated by any other solution. To evaluate the statistical significance of solutions, a statistical approach based on the concept of attainment function is used [4].

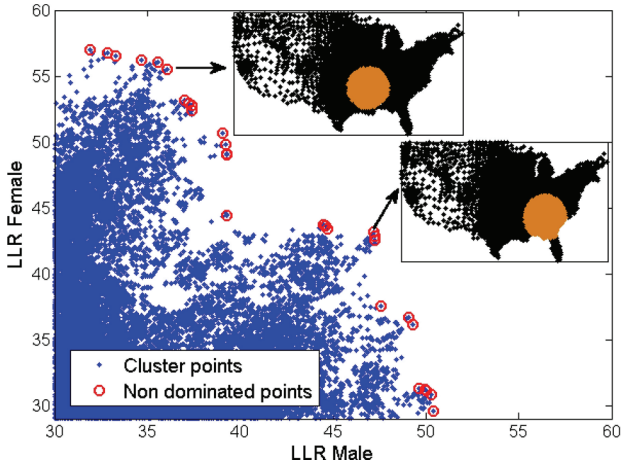
Results

The two datasets are standardized brain cancer mortality rates for male and female adults for each of the 3111 counties in the 48 contiguous states of the US, from 1986 to 1995 [5].

We run the circular scan and plot the (m(Zi),w(Zi)) points in the Cartesian plane, where m(Zi) and w(Zi) are the LLR for the zone Zi in the men's and women's brain cancer map, respectively, and i, i=1,...,N(r) is the set of all circular zones up to a radius r>0. The non-dominated set is inspected to observe possible correlations between the two maps regarding brain cancer clustering (Figure 1); e.g., the upper inset map has high LLR value on women's map, but not on men's; the inverse happens to the lower inset map. Other nondominated clusters in the middle have lower LLR values on both datasets. The first two examples have comparatively lower p-value (they belong to the two "knees" in the nondominated set), as computed using the attainment surfaces (not shown in the figure).

Conclusions

The multi-criteria multivariate approach has several advantages: (i) the representation of the evaluation function for each datastream is very clear, and does not suffer from an artificial, and possibly confusing mixture with the other datastream evaluations; (jj) it is possible to attribute, in a rigorous way, the statistical significance of each candidate cluster; (iii) it is possible to analyze and pick-up the best cluster solutions, as given naturally by the non-dominated set.



Part of the solution set in the LLR(male) X LLR(female) space of the male/female brain cancer datasets for the US counties map. Clusters are indicated by blue points, with the non-dominated solutions represented by small red circles. The inset maps depict the geographic location of the clusters found in the US counties map (yellow circles) for two sample non-dominated solutions.

Keywords

spatial scan statistic; Multi-criteria; attainment surface; Multiple data stream

Acknowledgments

The authors acknowledge the grants from CNPq and Capes.

References

[1] Kulldorff M, Mostashari F, Duczmal L, Yih K, Kleinman K, Platt R.(2007) Multivariate Scan Statistics for Disease Surveillance. Stat Med,26,1824-1833.

[2] Jonsson et al. Analysis of simultaneous space-time clusters of Campylobacter spp. in humans and in broiler flocks using a multiple dataset approach (2010). IJ Health Geogr,9:48

[3] Duczmal L, Cançado ALF, Takahashi RHC (2008) Geographic Delineation of Disease Clusters through multi-objective Optimization. J Comp Graph Stat,17:243-262.

[4] Cançado ALF, Duarte AR, Duczmal L, Ferreira SJ, Fonseca CM, Gontijo ECDM (2010). Penalized likelihood and multiobjective spatial scans for the detection and inference of irregular clusters. IJ Health Geogr,9:55.

[5] Fang Z, Kulldorff M, Gregorio DI (2004). Brain cancer mortality in the United States, 1986 to 1995: A geographic analysis. Neuro-Oncology 03-045, May 6.

*Luiz H. Duczmal
E-mail: duczmal@ufmg.br